
**ПРИБОРОСТРОЕНИЕ
ФИЗИКО-ХИМИЧЕСКОЙ БИОЛОГИИ**

УДК 543.426; 543.9

© Д. А. Белов, Ю. В. Белов

**КОМПЕНСАЦИЯ РАЗЛИЧИЯ ЭЛЕКТРОФОРЕТИЧЕСКОЙ
ПОДВИЖНОСТИ ФЛУОРЕСЦЕНТНО-МЕЧЕНЫХ
ФРАГМЕНТОВ ДНК ПРИ ГЕНЕТИЧЕСКИХ АНАЛИЗАХ**

В статье предложен способ компенсации различия электрофоретической подвижности флуоресцентно-меченых фрагментов ДНК при определении нуклеотидной последовательности. Определены неравномерности исходной последовательности пиков. Выполнена оптимизация компенсирующих параметров для каждого цветового канала. Оценены результаты компенсации различия электрофоретической подвижности флуоресцентно-меченых фрагментов ДНК, при этом показано, что погрешности определения временного положения пиков значительно уменьшаются и имеют случайный характер. На конкретном примере выполнена оптимизация параметров временного сдвига пиков для каждого цветового канала и построены уточненные графики в диапазоне до 300 нуклеотидов. С целью расширения диапазона применимости предложенного способа можно использовать калибровку и линеаризацию горизонтальной оси. Предложенный способ обеспечивает надежное определение пропущенных и "лишних" пиков.

Кл. сл.: ДНК, генетический анализатор, флуоресцентная детекция

ВВЕДЕНИЕ

Определение нуклеотидной последовательности (секвенирование ДНК) в генетическом анализаторе выполняется путем разделения фрагментов ДНК в капилляре под действием электрического поля. Четыре последовательности пиков, соответствующие нуклеотидам А, С, G и Т, регистрируются на выходе одноименных цветовых каналов флуоресцентного детектора. На экспериментальных "сырых" графиках наблюдается неравномерное распределение соседних пиков. Причина этого явления — различие электрофоретической подвижности флуоресцентно-меченых фрагментов ДНК [1–5.] Для определения нуклеотидной последовательности эти графики объединяются с использованием общего временного масштаба. При частичном перекрытии соседних пиков и значительных случайных ошибках вычисления временного положения пиков возможны ошибки при определении истинной нуклеотидной последовательности.

В статье [6] приведен способ вычисления аппроксимирующей функции, представляющей монотонную аналитическую зависимость базового временного интервала T_6 от номера пика в виде полинома 3-го порядка. С помощью базовой последовательности пиков с интервалом T_6 можно оценить неравномерность распределения соседних пиков.

В настоящей статье с целью значительного

уменьшения погрешности определения временного положения пиков предложен способ компенсации различия электрофоретической подвижности флуоресцентно-меченых фрагментов ДНК.

ПОСЛЕДОВАТЕЛЬНОСТЬ ВЫЧИСЛЕНИЙ

Для компенсации различия электрофоретической подвижности флуоресцентно-меченых фрагментов ДНК предлагается следующая последовательность вычислений в среде Excel:

- 1) получение экспериментальных данных разделения фрагментов секвенной смеси и определение времен выхода пиков;
- 2) оценка неравномерности последовательности пиков путем построения аппроксимирующей функции, представляющей зависимость базового временного интервала T_6 от времени (номера пика) и позволяющей выразить неравномерность интервалов между пиками в базовых относительных единицах (б.о.е.);
- 3) введение и оптимизация параметров временного сдвига пиков для каждого цветового канала;
- 4) получение обработанных данных и построение уточненных графиков;
- 5) оценка результатов компенсации различия электрофоретической подвижности флуоресцентно-меченых фрагментов ДНК.

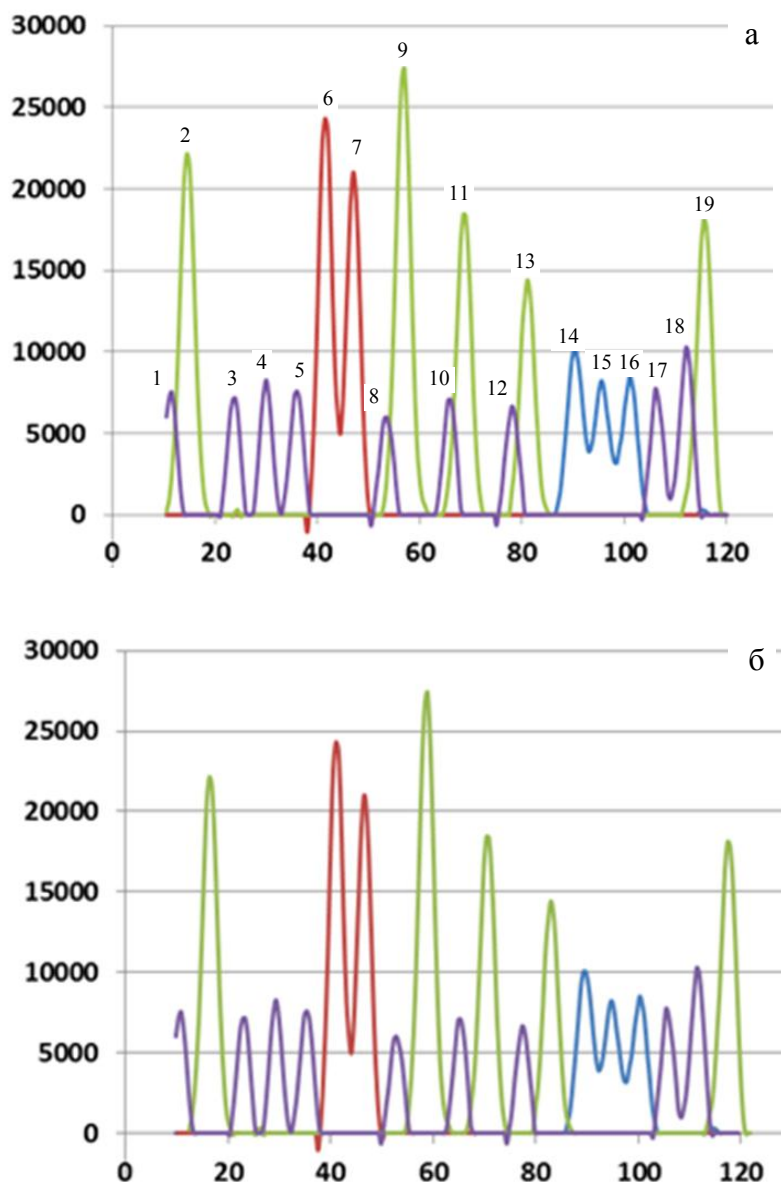


Рис. 1. Вид участка пиков, отражающих результаты разделения фрагментов секвенной смеси.

а — экспериментальные данные; б — результаты компенсации различия электрофоретической подвижности

По горизонтали — приращение времени в секундах, по вертикали — значение интенсивности флуоресценции в относительных единицах (о.е.) в четырех цветовых каналах: пики 1, 3, 4, 5, 8, 10, 12, 17, 18 — канал 1; пики 2, 9, 11, 13, 19 — канал 2; пики 6, 7 — канал 3; пики 14, 15, 16 — канал 4

РАЗДЕЛЕНИЕ ФРАГМЕНТОВ СЕКВЕННОЙ СМЕСИ

В качестве примера использованы экспериментальные данные разделения фрагментов секвенной смеси плазмидной ДНК, приведенные в статье [6]. При определении времен выхода пиков в четырех цветовых каналах флуоресцентного детектора использован программный модуль анализатора ДНК АЛ [7–9].

На рис. 1, а, показан небольшой участок экспериментальных данных, отражающих результаты разделения 19 фрагментов секвенной смеси.

На графиках рис. 1, а, видно, что некоторые со-

седние пики (1 и 2, 8 и 9, 10 и 11, 12 и 13, 18 и 19), соответствующие фрагментам ДНК, окрашенным разными красителями, в значительной степени взаимно перекрываются.

ОЦЕНКА НЕРАВНОМЕРНОСТИ ПОСЛЕДОВАТЕЛЬНОСТИ ПИКОВ

В табл. 1 приведены результаты вычисления аппроксимирующей функции, приведенные ранее в более полном составе в статье [6]. В таблице воспроизведен образ электронной таблицы Excel, использованной для вычислений.

Табл. 1. Вычисление отличий временных интервалов между соседними пиками от базового временного интервала

A	B	C	E	F	M	J
1906	1	G	1900.80	5.85	0.62	0.28
1915.5	2	C	1906.65	5.85	0.20	
1922.5	3	T	1912.50	5.86	-0.23	
1927	4	C	1918.36	5.86	-0.66	
1929	5	G	1924.21	5.86	0.11	
1935.5	6	G	1930.07	5.86	0.88	
1946.5	7	T	1935.93	5.86	-0.06	
1952	8	A	1941.79	5.86	-0.40	

В столбце А табл. 1 приведены положения центров пиков (время выхода в секундах), в столбцах В и С присвоены номера пиков и буквенные обозначения А, С, G и Т, соответствующие каждому конечному нуклеотиду фрагмента и присоединенному к нему красителю.

В столбце F вычислены базовые временные интервалы T_6 между соседними пиками (в секундах) на основе аппроксимирующей функции в столбце E.

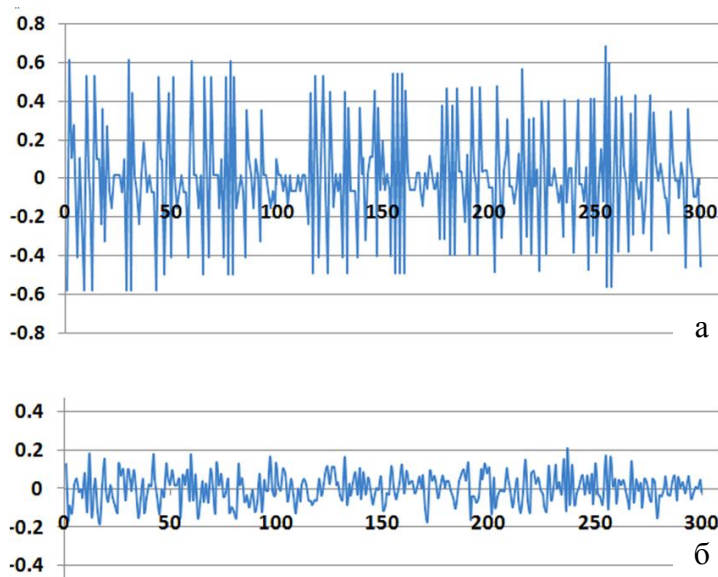
В столбце M приведены величины D_N — отличия временных интервалов между соседними пиками от базового временного интервала в базовых относительных единицах (б.о.е.). Величины D_N могут служить мерой неравномерности последовательности пиков, их можно рассматривать как сумму систематической и случайной составляю-

щих погрешности измерения взаимного положения пиков.

В ячейке J вычислено стандартное отклонение (значение, приблизительно равное СКО). Использовался оператор

$$J = \text{СТАНДОТКЛОН}(M51:M350) = 0.28 \text{ (б.о.е.)}.$$

Результаты вычислений величин D_N (столбец M) на участке 300 нуклеотидов в графическом виде представлены на рис. 2, а. На основе анализа результатов вычислений величин D_N можно предположить, что систематическая составляющая погрешности преобладает и ее можно компенсировать введением в дальнейший расчет соответствующих параметров.

**Рис 2.** Отличия временных интервалов между соседними пиками от базовых временных интервалов (D_N).

а — экспериментальные данные; б — компенсированные результаты. Вертикальная ось — величины D_N (б.о.е.). Горизонтальная ось — номера пиков. Для наглядности точки графика соединены прямыми линиями

ЗАДАНИЕ И ОПТИМИЗАЦИЯ КОМПЕНСИРУЮЩИХ ПАРАМЕТРОВ ДЛЯ КАЖДОГО ЦВЕТОВОГО КАНАЛА

В качестве компенсирующих параметров (поправок) предлагается использовать индивидуальный временной сдвиг пиков в каждом световом канале.

Результаты дальнейших вычислений приведены в табл. 2. Столбцы А, В и С табл. 2 заимствованы из табл. 1, столбец F табл. 1, который также будет использоваться, в табл. 2 не показан.

Поправки для каждого цветового канала в табл. 2 задаются в столбце L: L1 — для красителя А; L2 — для красителя С; L3 — для красителя G; L4 — для красителя Т. В начале расчета эти поправки принимаются, равными нулю.

В столбцах М, N, О и Р поправки прибавляются к временам выхода пиков в столбце А соответствующих цветовых каналов А, С, G и Т. В столбце Q получается компенсированная последовательность пиков: $Q = M + N + O + P$.

В столбце R вычислены компенсированные временные интервалы между соседними пиками (в секундах): $R_n = Q_{n+1} - Q_n$.

В столбце S вычислены погрешности — отличия компенсированных временных интервалов между соседними пиками R от базовых временных интервалов F из табл. 1 (в секундах):

$$S = R - F.$$

В ячейке T1 — сумма квадратов погрешностей

$$T1 = \text{СУММ}(S51^2:S350^2).$$

Величины в столбце L уточняются с помощью метода наименьших квадратов (минимум величины T1) и метода последовательного приближения в меню "Данные\Анализ\Поиск решения", изменяя величины поправок для каждого цветового канала в столбце L.

В столбце U вычислены величины относительных погрешностей взаимного положения компенсированных соседних пиков D_N . Значение ячейки U1 вычислено по формуле:

$$U1 = S1/F1,$$

где F1 — базовый временной интервал в первой строке табл. 1.

В ячейке T2 вычислено стандартное отклонение:

$$T2 = \text{СТАНДОТКЛОН}(U51:U350) = 0.08 \text{ (б.о.е.)}.$$

Результаты компенсации различия электрофоретической подвижности (величины D_N в столбце U) приведены на рис. 2, б.

ОЦЕНКА РЕЗУЛЬТАТОВ КОМПЕНСАЦИИ РАЗЛИЧИЯ ЭЛЕКТРОФОРЕТИЧЕСКОЙ ПОДВИЖНОСТИ ФЛЮОРЕСЦЕНТНО-МЕЧЕНЫХ ФРАГМЕНТОВ ДНК

В результате компенсации различия электрофоретической подвижности стали визуально значительно более равномерными расстояния между пиками на рис. 1, б, по сравнению с экспериментальными данными, приведенными на рис. 1, а.

Табл. 2. Вычисление погрешности взаимного положения соседних пиков при компенсации влияния различия электрофоретической подвижности флуоресцентно-меченых фрагментов ДНК

A	B	C	L	M	N	O	P	Q	R	S	T	U
1906	1	G	-0.79	0	0.00	1907.88	0.00	1907.88	7.16	1.31	61.8	0.22
1915.5	2	C	-0.46	0	1915.04	0.00	0.00	1915.04	6.81	0.95	0.08	0.16
1922.5	3	T	1.88	0	0.00	0.00	1921.84	1921.84	4.69	-1.16	—	-0.20
1927	4	C	-0.66	0	1926.54	0.00	0.00	1926.54	4.34	-1.52	—	-0.26
1929	5	G	—	0	0.00	1930.88	0.00	1930.88	6.50	0.64	—	0.11
1935.5	6	G	—	0.00	0.00	1937.38	0.00	1937.38	8.47	2.61	—	0.45
1946.5	7	T	—	0.00	0.00	0.00	1945.84	1945.84	5.36	-0.50	—	-0.08
1952	8	A	—	1951.21	0.00	0.00	0.00	1951.21	3.83	-2.03	—	-0.35

Можно сравнить величины D_N в столбце М табл. 1 и в столбце U табл. 2, а также соответствующие этим величинам графики на рис. 2, а, и рис. 2, б. По результатам сравнения можно сделать следующий вывод: после компенсации различия электрофоретической подвижности флуоресцентно-меченых фрагментов ДНК значительно уменьшаются максимальные погрешности и стандартное отклонение (примерно в 3–3.5 раза).

После компенсации погрешности измерения взаимного положения пиков D_N , приведенные на рис. 2, б, имеют случайный характер.

Уточненные графики на рис. 2 построены в диапазоне до 300 нуклеотидов. В этом диапазоне величина базового временного интервала T_6 изменяется незначительно. С целью расширения диапазона применимости предложенного способа можно использовать калибровку и линеаризацию горизонтальной оси с помощью известного размерного стандарта 1200 LIZ [10].

В случае, если относительные погрешности взаимного положения компенсированных соседних пиков превосходят по абсолютной величине значение 0.3 б.о.е., можно утверждать, что это признак пропущенного или "лишнего" пиков.

ЗАКЛЮЧЕНИЕ

1. Предложен способ компенсации различия электрофоретической подвижности флуоресцентно-меченых фрагментов ДНК при определении нуклеотидной последовательности.

2. На конкретном примере выполнена оптимизация параметров временного сдвига пиков для каждого цветового канала и построены уточненные графики в диапазоне до 300 нуклеотидов. В этом диапазоне величина базового временного интервала T_6 изменяется незначительно. С целью расширения диапазона применимости предложенного способа можно использовать калибровку и линеаризацию горизонтальной оси.

3. Предложенный способ обеспечивает надежное определение пропущенных и "лишних" пиков. Для этого используются в качестве критерия отличия компенсированных временных интервалов между соседними пиками от базовых временных интервалов.

4. Оценены результаты компенсации различия электрофоретической подвижности флуоресцентно-меченых фрагментов ДНК: погрешности определения временного положения пиков значительно уменьшаются и имеют случайный характер.

СПИСОК ЛИТЕРАТУРЫ

1. Будилов А. Методы расшифровки нуклеотидной последовательности фрагментов ДНК. URL: http://molbiol.ru/protocol/13_03.html.
2. Tu O., Mnott T., Marsh M. et al. The influence of fluorescent dye structure on the electrophoretic mobility of end-labeled DNA // *Nucleic Acids Research*. 1998. Vol. 26, no. 11. P. 2797–2802.
3. Ju J., Glazerl A.N., Mathies R.A. Cassette labeling for facile construction of energy transfer fluorescent primers // *Nucleic Acids Research*. 1996. Vol. 24, no. 6. P. 1144–1148.
4. Лунина Н.Л. Система обработки нуклеотидных последовательностей HEID. URL: http://www.impb.ru/pdf/NL_1984_1r.pdf.
5. Флуоресцентные красители. ЗАО "Синтол". URL: <http://www.syntol.ru/infoflu.htm>.
6. Алексеев Я.И., Белов Д.А., Белов Ю.В., Курочкин В.Е. Исследование погрешностей оцифровки пиков генетического анализатора // *Научное приборостроение*. 2014. Т. 24, № 2. С. 79–85. URL: <http://213.170.69.26/mag/2014/full2/Art10.pdf>.
7. Леонтьев И.А. Обработка данных в задачах электрофореза // *Научное приборостроение*. 2003. Т. 13, № 2. С. 96–99.
8. Леонтьев И.А. Обсчет пиков в задачах электрофореза // *Научное приборостроение*. 2004. Т. 14, № 1. С. 94–96.
9. Алексеев Я.И., Белов Ю.В., Малюченко О.П. и др. Генетический анализатор для фрагментного анализа ДНК // *Научное приборостроение*. 2012. Т. 22, № 4. С. 86–92. URL: <http://213.170.69.26/mag/2012/full4/Art12.pdf>.
10. URL: http://www6.appliedbiosystems.com/support/software/genescan_sizestandards/GS1200LIZ_Size_Standard_Definition_import_instructions.pdf.

*Институт аналитического приборостроения РАН,
г. Санкт-Петербург*

Контакты: Белов Дмитрий Анатольевич,
onoff_10@mail.ru

Материал поступил в редакцию: 22.04.2016

FLUORESCENTLY MARKED DNA FRAGMENTS ELECTROPHORETIC MOBILITY COMPENSATION IN GENETIC ANALYSIS

D. A. Belov, Yu. V. Belov

Institute for Analytical Instrumentation of RAS, Saint-Petersburg, Russia

The main factor of adjacent peaks uneven distribution in the experimental graphs during the nucleotide sequences determination which are recorded on four color channels outputs of genetic analyzer fluorescence detector is discussed in the article. The reason of this is the difference in the electrophoretic mobility of fluorescently marked DNA fragments. DNA fragments electrophoretic mobility differences compensation method is suggested and detailed sequence of calculations is given. The experimental data of plasmid DNA fragment separation is taken as an example. The peak time values and the unevenness of the original sequence of peaks in the four color channel fluorescence detector are determined, fluorescently marked DNA fragments electrophoretic mobility differences compensation results are estimated. Graphs of small segments of peaks sequence reflecting the results of the fragments separation at compensating differences in electrophoretic mobility are shown. The peak time shift parameters optimization for each color channel is made and refined graphs are built in the range of up to 300 nucleotides. It is also shown that the errors in determining the peaks time position are reduced significantly and have a random character. Horizontal axis calibration and linearization can be used in order to expand the proposed method applicability. The proposed method provides a reliable determination of missing and "extra" peaks.

Keywords: DNA, the genetic analyzer, fluorescence detection

REFERENCES

1. Budilov A. *Metody rasshifrovki nukleotidnoj posledovatel'nosti fragmentov DNK* [Methods of interpretation of nucleotide sequence of fragments of DNA]. Available at: http://molbiol.ru/protocol/13_03.html (In Russ.).
2. Tu O., Mnott T., Marsh M., Bechtol K., Harris D., Barker D. and Bashkin J. The influence of fluorescent dye structure on the electrophoretic mobility of end-labeled DNA. *Nucleic Acids Research*, 1998, vol. 26, no. 11, pp. 2797–2802. Doi: 10.1093/nar/26.11.2797.
3. Ju J., Glazer A.N., Mathies R.A. Cassette labeling for facile construction of energy transfer fluorescent primers. *Nucleic Acids Research*, 1996, vol. 24, no. 6, pp. 1144–1148. Doi: 10.1093/nar/24.6.1144.
4. Lunina N.L. *Sistema obrabotki nukleotidnyh posledovatel'nostej HEID* [System of processing of nucleotide sequences of HEID]. Available at: http://www.impb.ru/pdf/NL_1984_1r.pdf (In Russ.).
5. *Fluorescentnye krasiteli* [Fluorescent dyes]. ZAO "Syntol". Available at: <http://www.syntol.ru/infoflu.htm> (In Russ.).
6. Alekseev Ya.I., Belov D.A., Belov Yu.V., Kurochkin V.E. [Research of the genetic analyzer digitization peaks errors]. *Nauchnoe Priborostroenie* [Scientific Instrumentation], 2014, vol. 24, no. 2, pp. 79–85. Available at: <http://213.170.69.26/mag/2014/full2/Art10.pdf> (In Russ.).
7. Leontyev I.A. [Data analysis in capillary electrophoresis applications]. *Nauchnoe Priborostroenie* [Scientific Instrumentation], 2003, vol. 13, no. 2, pp. 96–99 (In Russ.).
8. Leontyev I.A. [Peak calculations in electrophoresis problems]. *Nauchnoe Priborostroenie* [Scientific Instrumentation], 2004, vol. 14, no. 1, pp. 94–96 (In Russ.).
9. Alekseev Ya.I., Belov Yu.V., Malyuchenko O.P., Monahova Yu.A., Natyrov A.N., Orekhov V.A., Kononov S.V., Kurochkin V.E., Petrov A.I. [Genetic analyser for DNA fragment analysis]. *Nauchnoe Priborostroenie* [Scientific Instrumentation], 2012, vol. 22, no. 4, pp. 86–92. Available at: <http://213.170.69.26/mag/2012/full4/Art12.pdf> (In Russ.).
10. Instructions for importing GeneScan™ 1200 LIZ® Size Standard definition into GeneMapper® Software. Available at: http://www6.appliedbiosystems.com/support/software/genescan_sizestandards/GS1200LIZ_Size_Standard_Definition_import_instructions.pdf.

Contacts: Belov Dmitriy Anatol'evich,
onoff_10@mail.ru

Article received in edition: 22.04.2016